PB-216 755

CONFIDENCE LIMITS FOR AN UNKNOWN CONCENTRATION ESTIMATED FROM LINEAR CALIBRATION

Eugene K. Harris

General Electric Company
Cincinnati, Ohio

July 1958

DISTRIBUTED BY:

NTIS

National Technical Information Service
U. S. DEPARTMENT OF COMMERCE
5285 Port Royal Road, Springfield Va. 22151

LIBRARY
SANITARY ENGINEERING CENTER
CINCINNATI, OHIO

PB 216 755

AGT - TECHNICAL INFORMATION SERIES

R58AGT494

CONFIDENCE LIMITS FOR AN UNKNOWN CONCENTRATION ESTIMATED FROM LINEAR CALIBRATION

EUGENE K. HARRIS

AIRCRAFT GAS TURBINE DIVISION
FLIGHT PROPULSION LABORATORY DEPARTMENT
GENERAL ELECTRIC
CINCINNATI 15, OHIO

Reproduced by
NATIONAL TECHNICAL
INFORMATION SERVICE
U S Department of Commerce
Springfield VA 22151

GENERAL ELECTRIC

TECHNICAL INFORMATION SERIES

Title Page

AUTHOR	SUBJECT CLASSIFICATION	NO.
Eugene K. Harris		R58AGT494
		DATE: July, 1958

TITLE

Confidence Limits for An Unknown Concentration Estimated from Linear Calibration

ABSTRACT

G.E. CLASS	REPRODUCIBLE COPY FILED AT	NO. PAGES
I	Central Library, AGT Building 305	4
GOV. CLASS. None		

CONCLUSIONS

By cutting out this rectangle and folding on the center line, the above information can be fitted into a standard card file.

For list of contents — drawings, photos, etc. and for distribution see next page (GT 2063-B1).

INFORMATION PREPARED FOR Evendale Computations - Testing Operation, FPLD

TESTS MADE BY

COUNTERSIGNED OPERATION Evendale Computations, Testing Oper.

DIVISION AGT LOCATION Building 305, Evendale, Ohio

GT 2063-A1 (10-52)

Ia

R58AGT494

CONFIDENCE LIMITS FOR AN UNKNOWN CONCENTRATION ESTIMATED FROM LINEAR CALIBRATION

by

Eugene K. Harris, Ph.D.

July, 1958

Flight Propulsion Laboratory Department
General Electric Company
Cincinnati, Ohio

Ib

GENERAL ELECTRIC COMPANY

TECHNICAL INFORMATION SERIES

NO. R58AGT494

Initial study of an analytical method for determining the amount of some substance frequently leads to a calibration curve on which the response of a test indicator are plotted against carefully measured amounts of the substance.

In practice, however, such a curve is used to estimate an unknown amount of substance from an observed value of the indicator variable - exactly converse to the way in which it is prepared. Berkson(1) has shown that this estimation is valid when the relationship between substance (x) and indicator (y) is linear, provided that in obtaining the calibration line an attempt is made to measure x to certain predetermined amounts. The estimation of x from y is not invalidated by the fact that the original measurement of x may have been subject to error. The purpose of this note is to call attention to confidence limits available for such an estimate.

Assuming the linear relation,

$$\text{True } y = \alpha + \beta x$$

least squares estimates of α and β may be obtained from the calibration data by the usual formulae:

$$b = \text{Estimate of } \beta = \frac{\Sigma (x-\bar{x})(y-\bar{y})}{\Sigma (x-\bar{x})^2}, \text{ and}$$

$$a = \text{Estimate of } \alpha = \bar{y} - b\bar{x}$$

where $\bar{x},\bar{y}$ are the arithmetic means of the x's and y's, respectively, and there are n pairs of x,y values. If $m > 1$ determination of y are made at a given x-point, there are actually m pairs at that point, and the value of x must be repeated m times in the computations.

In subsequent applications of the calibration curve, one or more (say, m') replicate determination of y are made on an unknown x, the mean, $\bar{y}_r$, is computed, and the estimate

$$\hat{x} = \text{Estimate of } x = \frac{\bar{y}_r - a}{b}$$

is obtained. (In practice, this estimate is usually taken directly from the graph of the calibration curve.)

To compute confidence limits for the unknown x, two assumptions must be made: (1) that individual determinations (y) at each x-point are normally distributed, and (2) that the precision of y, measured by its variance (or standard deviation) is constant for all x. Under these assumptions, exact confidence limits for x may be obtained from a formula developed by Fieller(2) and also available in a paper by Geary(3). First, let

$$g = \frac{t^2 V(b)}{b^2}$$

where V denotes variance, and t is Student's "t" with n-2 degrees of freedom at a specified level of confidence. Then exact limits for x are given by the complicated expression

$$\frac{\hat{x} - g\dfrac{\text{Cov}\,[(\bar{y}_r - a), b]}{V(b)}}{1-g} \pm \frac{t}{b(1-g)} \sqrt{\begin{aligned} &V(\bar{y}_r - a) - 2\hat{x}\,\text{Cov}\,[(\bar{y}_r - a), b] + \hat{x}^2 V(b) \\ &- g\left\{V(\bar{y}_r - a) - \frac{\text{Cov}^2\,[(\bar{y}_r - a), b]}{V(b)}\right\} \end{aligned}} \qquad (1)$$

Property of General Electric Company

where Cov denotes covariance, and $V(\bar{y}_r - a) = V(\bar{y}_r) + V(a)$.

In most calibration experiments based on a linear relationship, the straight line will be fairly well defined by the observed points; hence, V(b)/b will be small, its square even smaller, and g practically zero. This leads to closely approximate limits given by the simpler expression

$$\hat{x} \pm \frac{t}{b}\sqrt{V(\bar{y}_r - a) - 2\hat{x}\,\mathrm{Cov}\,(\bar{y}_r - a),\, b) + \hat{x}^2\,V(b)} \qquad (2)$$

From the calibration experiment,

$$\text{Estimated } V(a) = \frac{s_y^2 \Sigma x^2}{n \Sigma(x-\bar{x})^2}, \qquad \text{Estimated } V(b) = \frac{s_y^2}{\Sigma(x-\bar{x})^2},$$

and $$\text{Estimated Cov}\,(\bar{y}_r - a)\, b) = -\text{Est. Cov}\,(a,b) = \frac{s_y^2\,\bar{x}}{\Sigma(x-\bar{x})^2},$$

where s_y^2, the estimated variance of a single determination of y, is computed from the formula

$$s_y^2 = \Sigma(y - a - bx)^2/n-2$$
$$= \frac{\Sigma y^2 - n\bar{y}^2 - b^2\Sigma(x-\bar{x})^2}{n-2}$$

Finally, Estimated $V(\bar{y}_r) = s_y^2/m'$, assuming that the precision of future determinations is the same as that observed during the calibration experiments.

Substituting these formulae into Equation (2), confidence limits for x are given by

$$\hat{x} \pm \frac{ts_y}{b}\sqrt{\frac{1}{m'} + \frac{1}{n} + \frac{(\hat{x}-\bar{x})^2}{\Sigma(x-\bar{x})^2}} \qquad (3)$$

This formula may be found in Bennett and Franklin(4).

The assumption of constant V(y) for all x is frequently not satisfied by the data in original units. Of course, this assumption can be tested only when more than one determination of y is available at each x. A standard test procedure is Bartlett's Chi-Square test(5). When the variance appears heterogeneous, the observations should be transformed to a scale on which the variance is approximately constant. A variance function frequently met in chemical experiments is V(y, at a given x) = a(ȳ, at that x)^2. In this case, the logarithms of the sample variances will plot linearly against the logarithms of the sample means of y. If the slope of this line is close to unity, a transformation to the square root is indicated; if close to two, the logarithmic scale should be chosen; if close to three, the reciprocal of the square root; if close to four, the reciprocal of y itself. Final results are always "re-transformed" into original units.

It is necessary that y on the transformed scale plot linearly against x or some function of x. Frequently, the transform of y will plot linearly against the same transform of x; in fact, such transformations may not only stabilize the variance of y but also convert a non-linear relationship to a

Property of General Electric Company

linear one. This is illustrated in the following example, and has been given some theoretical basis in recent work by Geary(6) indicating that the variance of y will not be constant when the calibration curve is non-linear.

Example: Table I presents data of Brown and Marsh(7) on the chromatographic assay of nicotinamide by absorption on paper strips.

TABLE I

Assay of Nicotinamide
by Absorption on Paper Strips (Brown and Marsh(8))

x (nicotinamide: g/100 ml)	log x	y (area of absorption zone: sq. in.)
.0800	-1.0969	2.00,2.05 2.12,1.96 Mean = 2.03; Variance = .00477 Variance log y = .0002149
.1600	-0.7959	3.35,3.36 3.41,3.55 Mean = 3.42; Variance = .00850 Variance log y = .0001345
.2400	-0.6198	4.35,4.29 4.39,4.72 Mean = 4.44; Variance = .03717 Variance log y = .0003421
.3200	-0.4948	5.50,6.05 5.92,5.80 Mean = 5.82; Variance = .05523 Variance log y = .0003141
.4000	-0.3979	6.18,6.37 6.75,6.26 Mean = 6.39; Variance = .06367 Variance log y = .0002859
	Σ log x = -13.6212 Mean log x = -0.68106 $\Sigma(\log x)^2$ = 10.495800	Σ log y = 12.2320 Mean log y = 0.6116 $\Sigma(\log y)^2$ = 8.126828

n = 20; Σ(log x log y) = -7.447767

The first thing to notice is that the variance of y in square inches increases with the mean; log variance plots more or less linearly against log mean with slope 2.4. The logarithmic transformation is indicated and produces a relatively stable variance, as may be judged from the values of variance log y given in Table I. Bartlett's test yields a quite insignificant Chi-Square of 0.77 with four degrees of freedom.

Property of General Electric Company

Furthermore, the authors have shown that the plot of log y against log x is linear, indicating the usefulness of the transformation. Least squares estimates of the slope and intercept of this line are b = 0.7244 and a = 1.1049, respectively. The estimated variance of log y, $s^2_{\log y}$, is 0.0003419.

Now, suppose that at some future time, a series of four replicate determinations on an unknown amount of nicotinamide (x) yields a mean absorption area of 5.02 square inches. Then,

$$\log \hat{x} = \frac{\log (5.02) - a}{b} = -0.5580,$$

whence the estimate of x is 0.2767 g/100 ml.

Confidence limits are first obtained for log x. At the 95 percent probability level, with eighteen degrees of freedom, t = 2.10. From Equation (3) and the statistics given at the bottom of Table I, 95 percent confidence limits for log x are -0.5886 and -0.5274. The corresponding limits for x, the unknown concentration of nicotinamide, are 0.2579 and 0.2969 g/100 ml.

LITERATURE CITED

(1) Berkson, J., J. Amer. Stat. Assoc. 45, 164 (1950).

(2) Fieller, E. C., Quart. Jour. Pharm., 17, 117 (1944).

(3) Geary, R. C., J. Roy. Stat. Soc., 93, 442 (1930).

(4) Bennett, C. A. and Franklin, N. L. "Statistical Analysis in Chemistry and the Chemical Industry". John Wiley and Sons, Inc. New York, 1954, P. 232.

(5) Snedecor, G. W., "Statistical Methods:, 4th ed., Ames, the Collegiate Press, Inc., 1946, P. 250.

(6) Geary, R. C., J. Amer. Stat. Assoc., 48, 94 (1953).

(7) Brown, J. A., and Marah, M. M., Anal. Chem., 24, 1952 (1952).

Property of General Electric Company

DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE

PUBLIC HEALTH SERVICE

DIRECTOR
Robert A. Taft Sanitary Engineering Center
4676 Columbia Parkway
Cincinnati 26, Ohio

June 18, 1958

Mr. Gregory P. Williams
2532 Highland Avenue
Cincinnati 19, Ohio

Dear Mr. Williams:

I hereby grant permission to the General Electric Company to publish the attached paper in the General Electric technical information series.

Yours very truly,

Eugene K. Harris, Ph.D.
Analytical Statistician

Attachment

.5